

Lending Club Case Study

Presented by:

- Nagesh Kumar Vashist
- Mayank Aditya



Agenda

- Problem Statement
- Data Load and Cleanup
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Conclusion

Problem Statement

Lending Club is a **consumer finance company** which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's decision:

- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
- If the applicant is **not likely to repay the loan**, i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

As a data analyst we need to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending at a higher interest rate

Data load and cleanup

Data Cleanup is a process to remove unwanted data or noise in the data and convert the data to the desired format so that data analysis can be done on the dataset.

Below are the steps we have done for data cleanup:

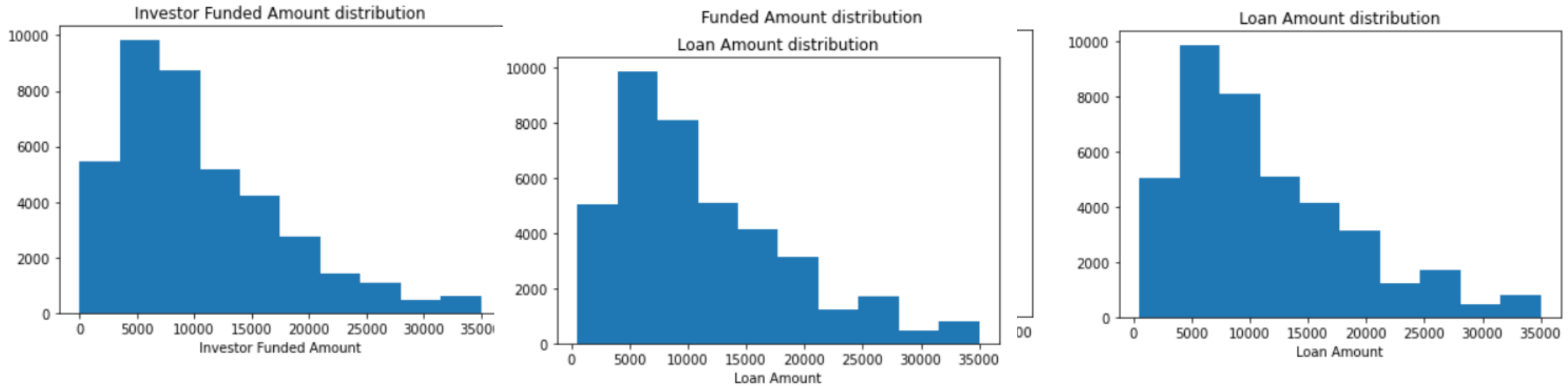
- Loaded loan.csv file with 39717 rows and 111 columns
- Removed all the empty columns
- Removed all the columns that are empty more than 70%
- Converted the column values to the expected format and data type
- Validated duplicate rows based on id column
- Created new columns for further analysis

Univariate Analysis

Done the analysis for important features as follows:

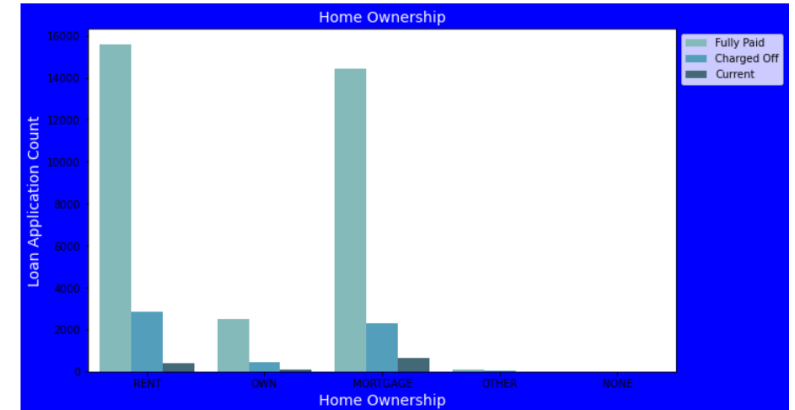
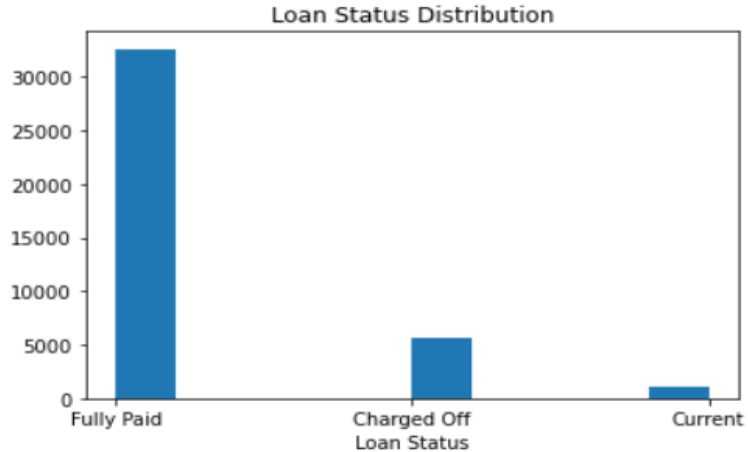
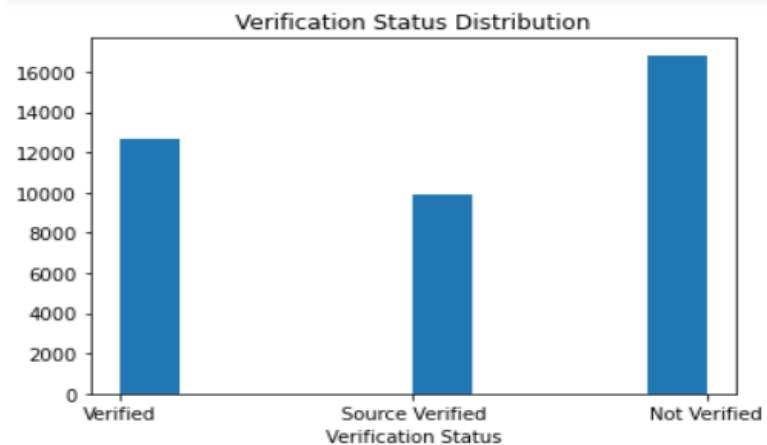
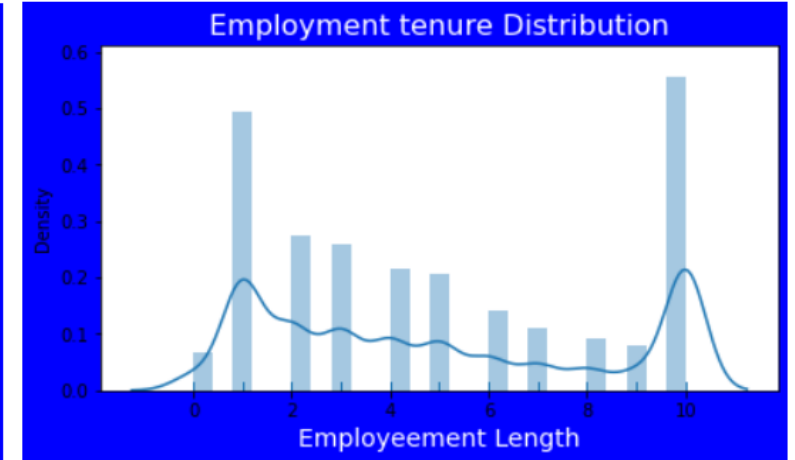
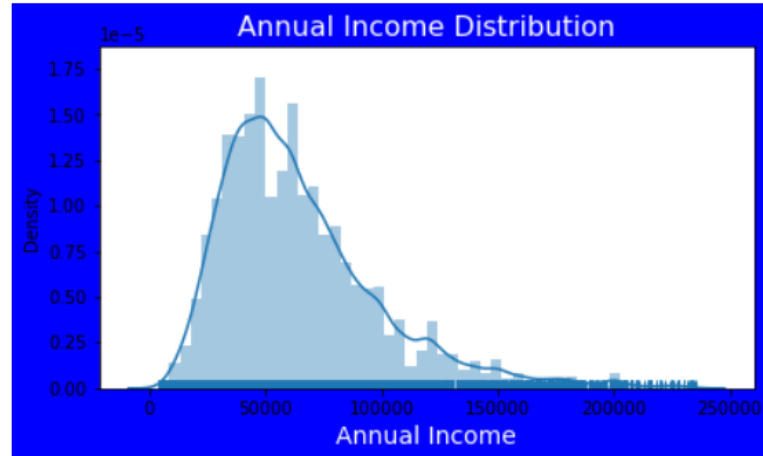
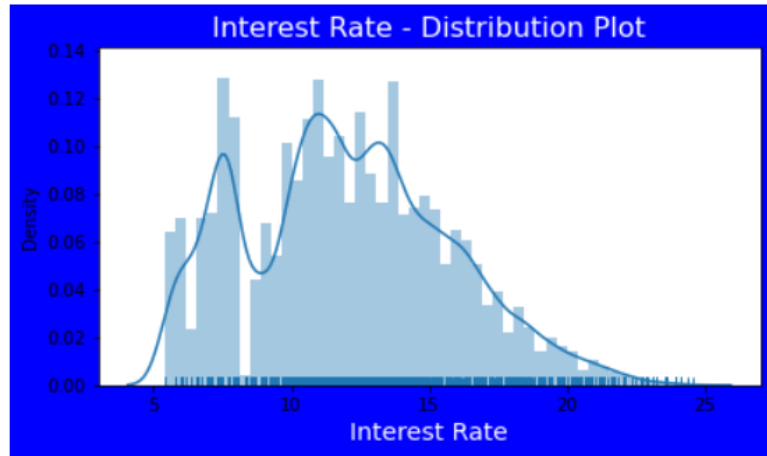
- Investor Funded Amount
- Interest Rate
- Annual Income
- Employment Tenure
- Loan Status
- Grades
- Verification Status
- Purpose of loan
- Home Ownership

Univariate Analysis of Loan amt



We can see that with above distribution plots Investor Funded Amount, Funded Amount and Loan Amount has similar distribution. For further analysis we will consider Investor Funded Amount

Few of the univariate plots

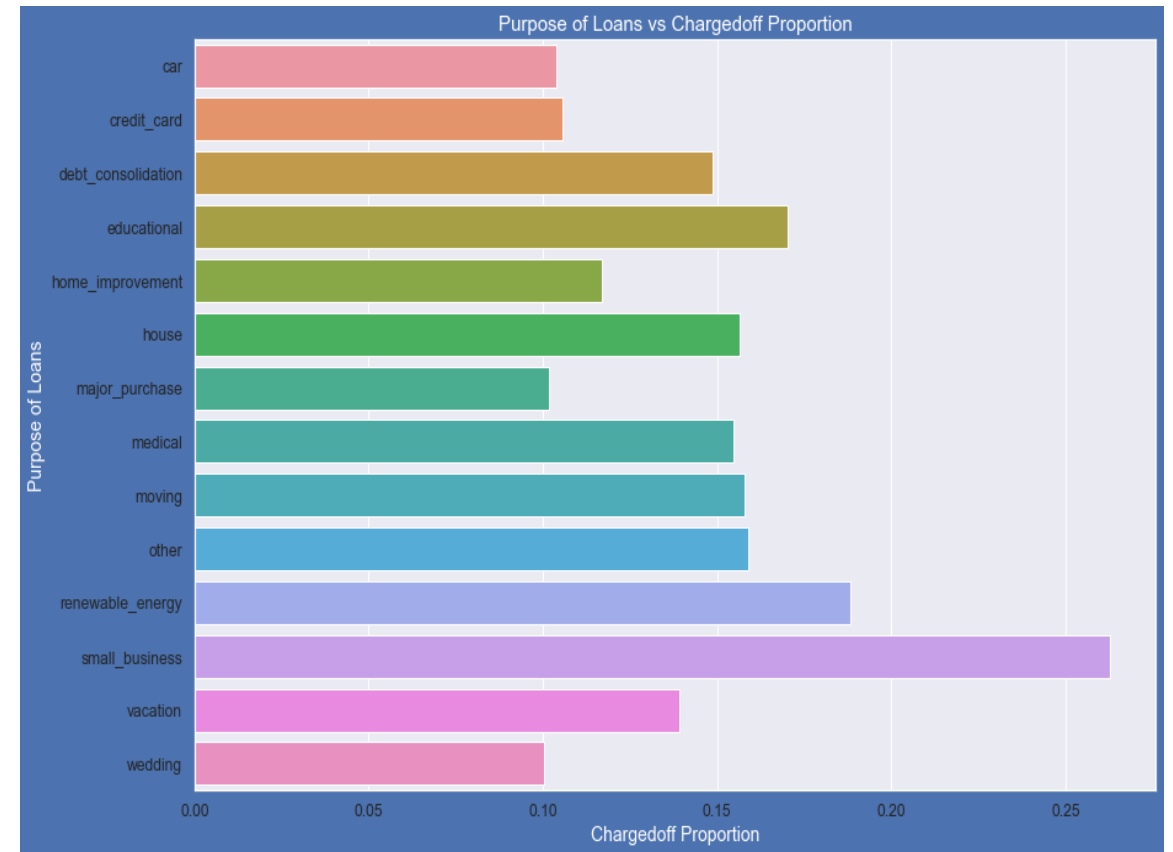
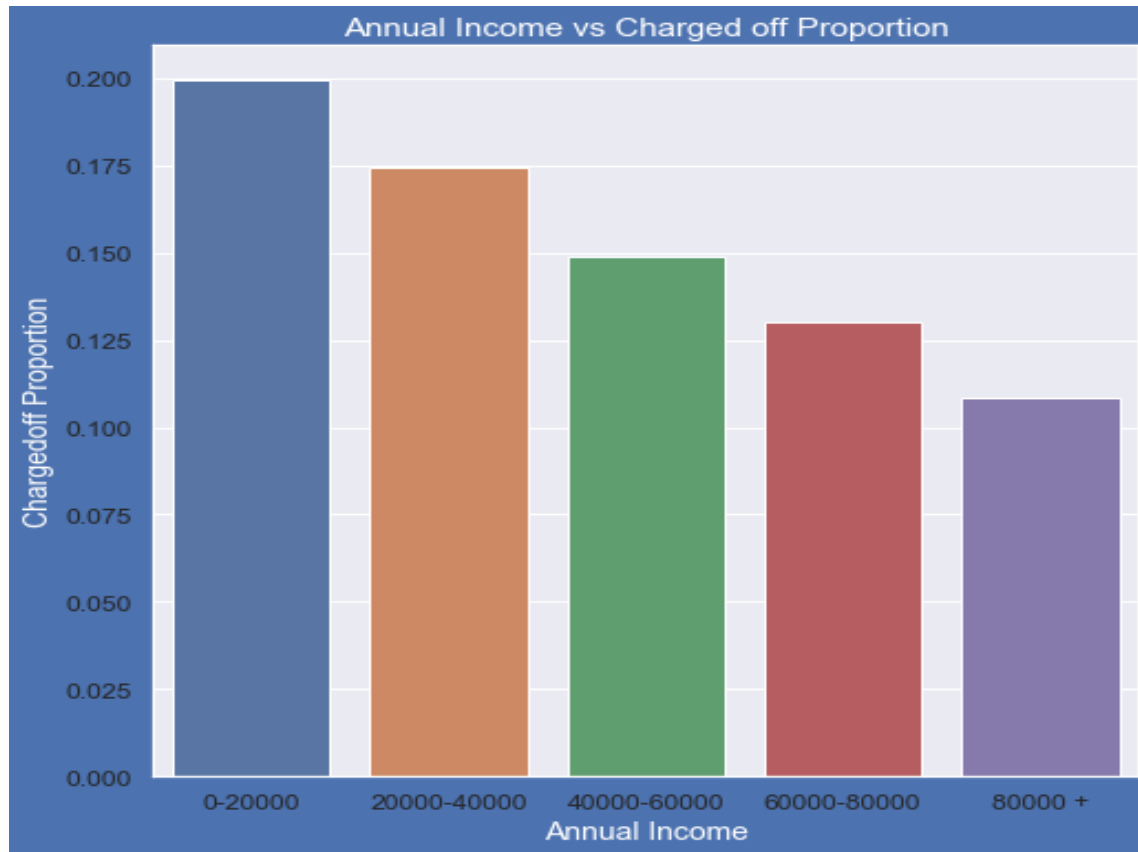


Bivariate Analysis

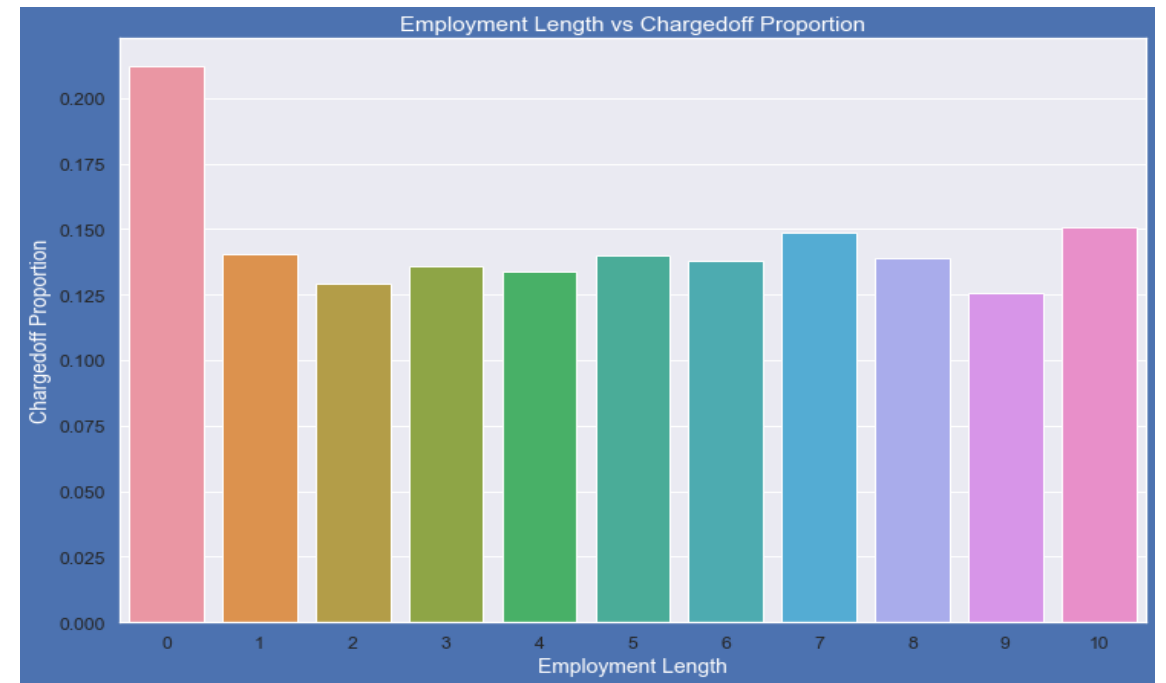
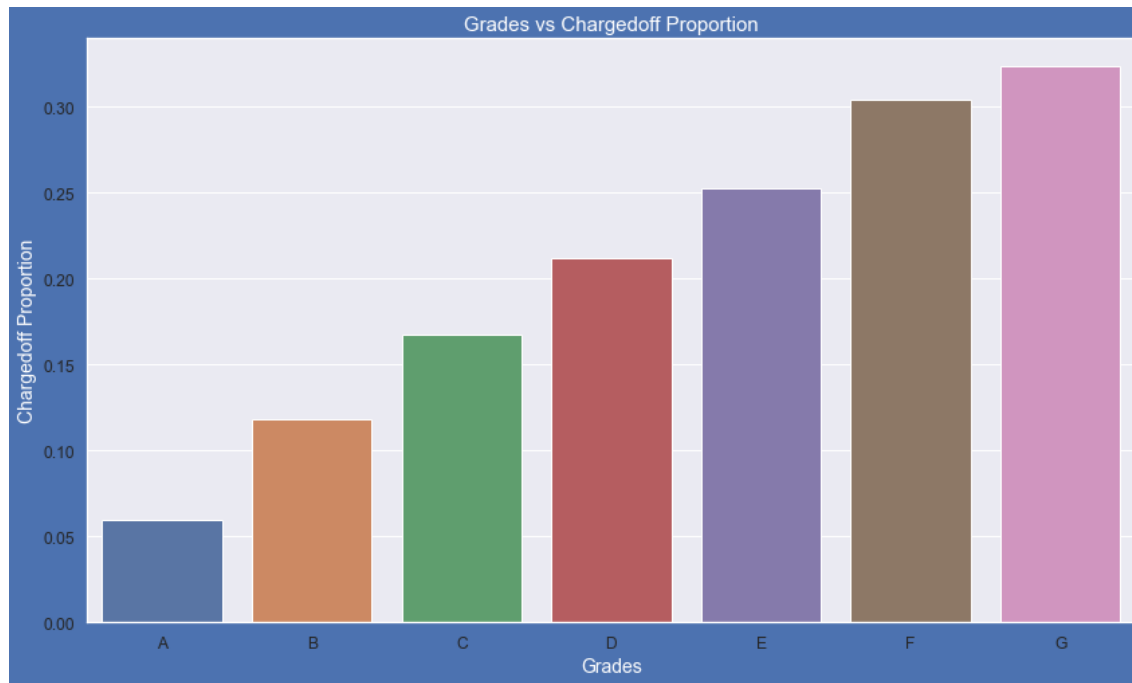
Done the analysis for important features as follows:

- Annual income and charged off proportion
- Purpose of loan and charged off proportion
- Grades and charged off proportion
- Interest rate and charged off proportion
- Employment length and charged off proportion
- State and charged off proportion
- Bankruptcies and charged off proportion

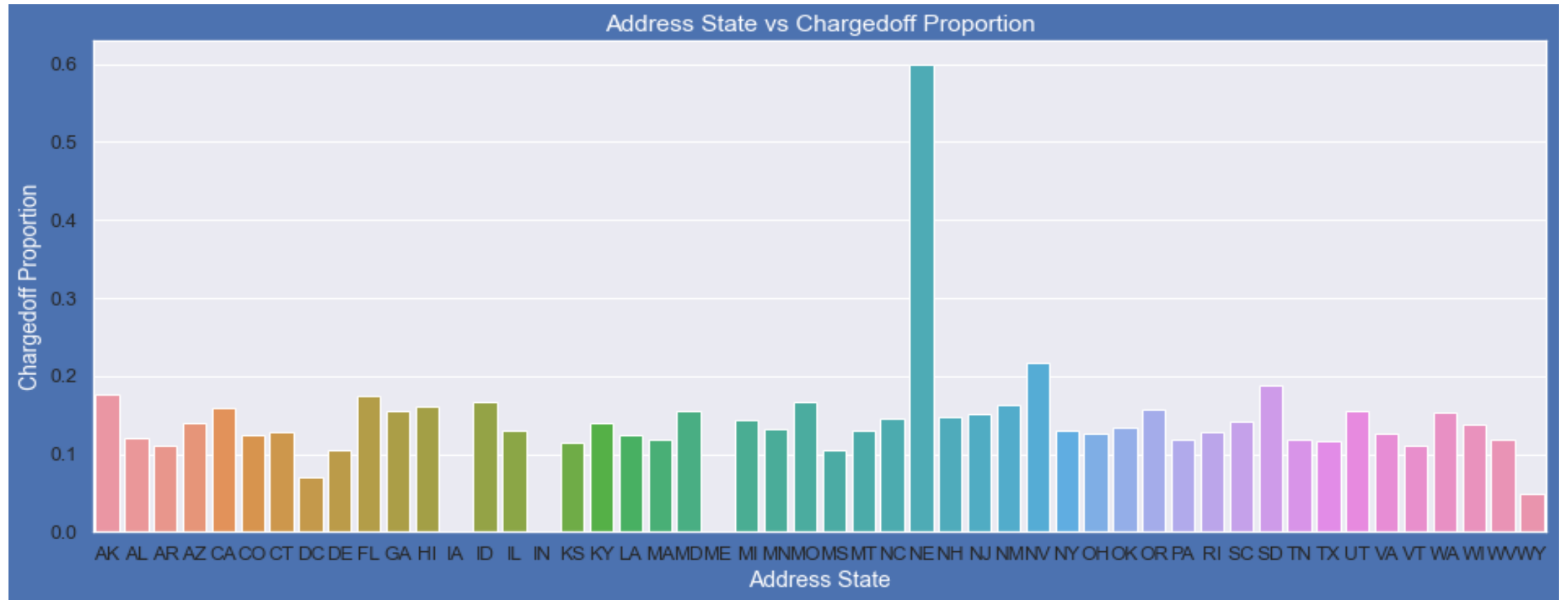
Few of the Bivariate plots



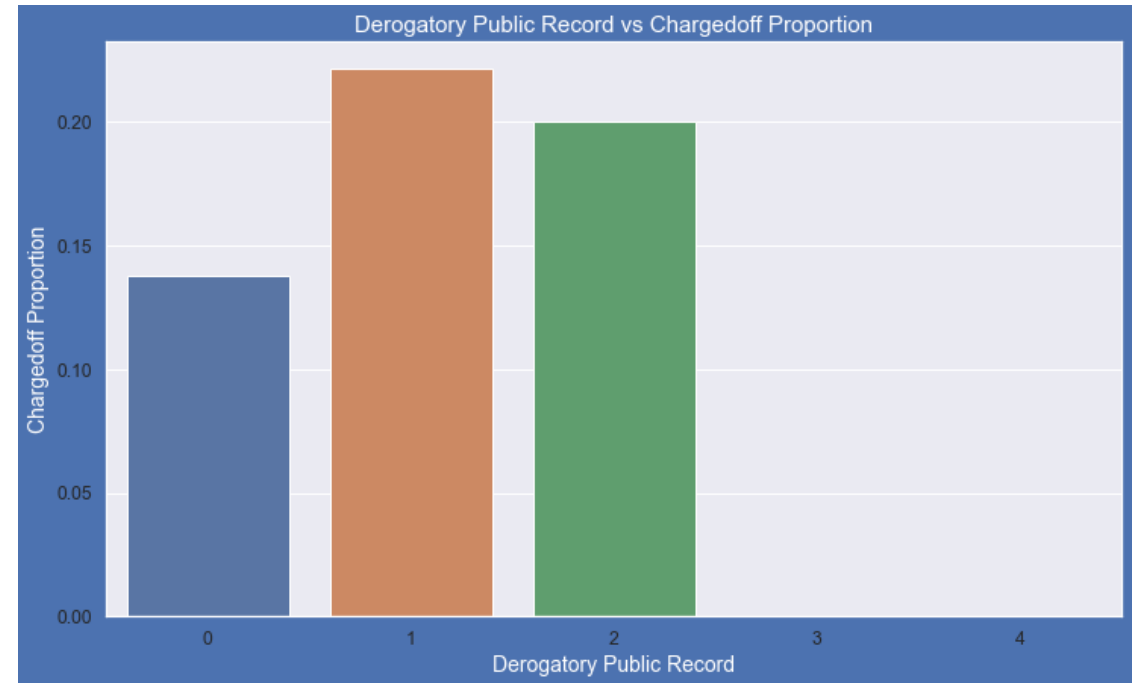
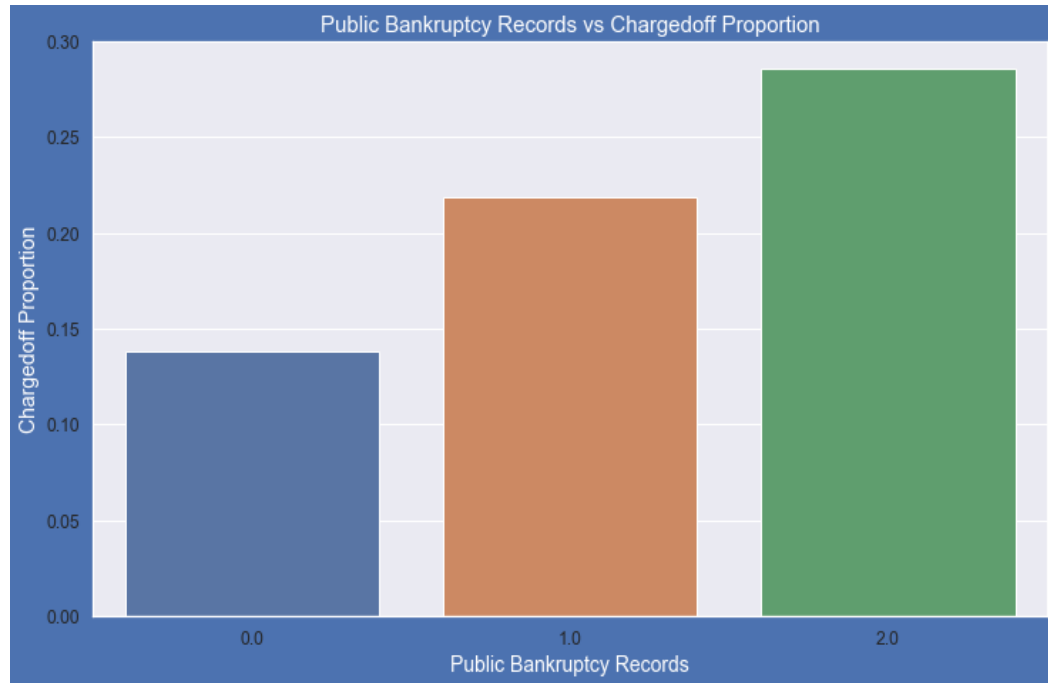
Few of the Bivariate plots



Few of the Bivariate plots



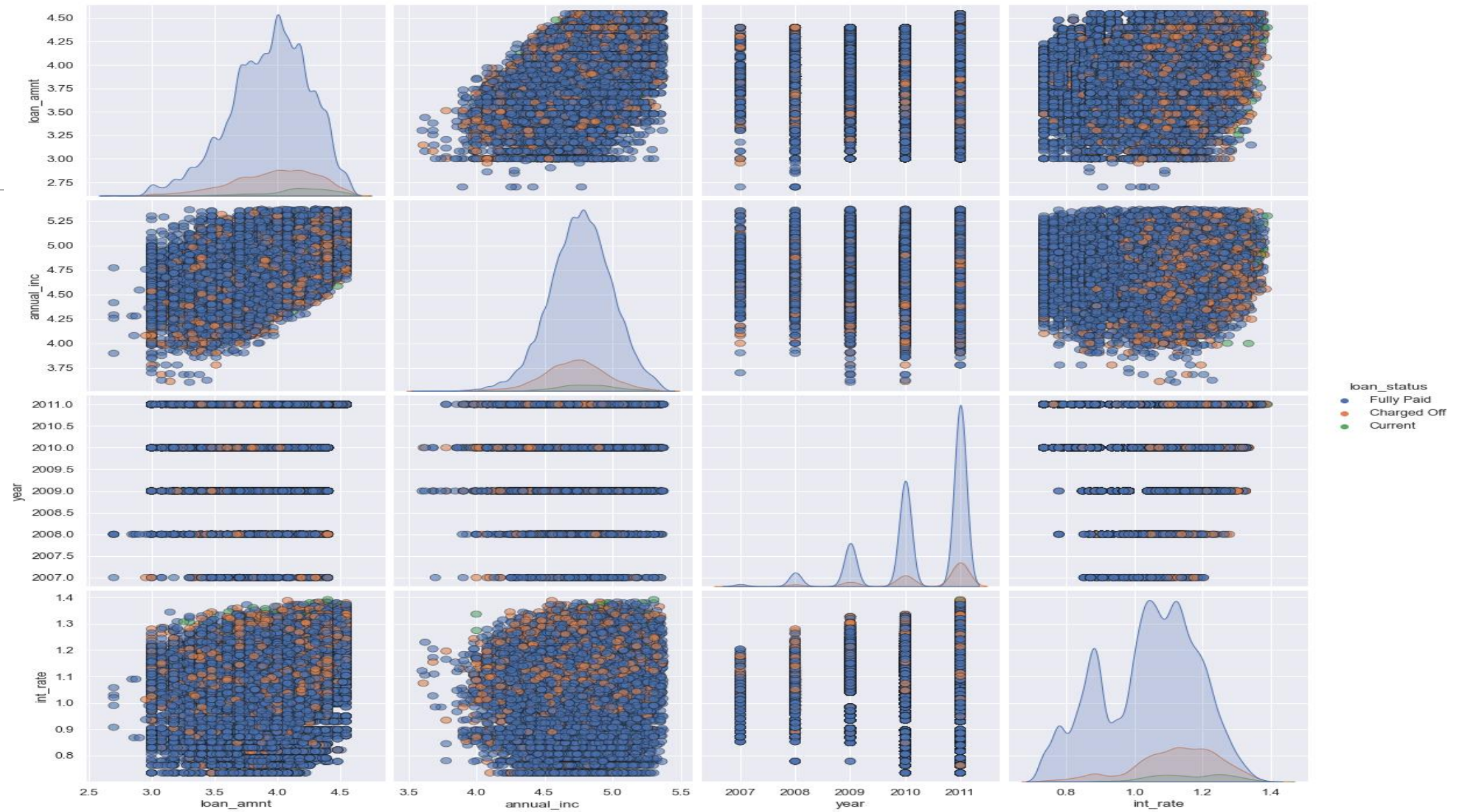
Few of the Bivariate plots



Multivariate Analysis

Plot following features differentiating with loan status

- loan_amnt
- annual_inc
- Year
- int_rate



Conclusion

- Higher the interest rate higher charged off ratio
- Higher the annual income higher the loan amount slightly
- Interest rate is increasing with loan amount increase this results in high charged off.
- Median,95th percentile,75th percentile of loan amount is highest for loan taken for small business purpose among all purposes.
- Debt consolidation is second and Credit card comes 3rd.
- Most of the loans issued for longer term had higher interest rates for repayment.
- Loans taken for small business purposes had to repay the loan with more interest rate as compared to other.
- Those who already have pub_rec_bankruptcies value 1, have charged off proportion higher than who have no pub_rec_bankruptcies.
- pub_rec_bankruptcies count 2 has even higher charged off proportion but those numbers are not significant to decide.
- states NE has very high chances of charged off but number of applications are too low to make any decisions.
- Those who are not working or have less than 1 year of work experience have high chances of getting charged off.
- sub Grades of "A" has very less chances of charged off.- sub Grades of "F" and "G" have very high chances of charged off.

*Thank
you!*